Supplementary Materials for

Helminth Infection Promotes Colonization Resistance via Type 2 Immunity

Deepshika Ramanan[1,2,*], Rowann Bowcutt[3]*, Soo Ching Lee[4], Mei San Tang[3], Zachary D. Kurtz[3], Yi Ding[5], Kenya Honda[6,7], William C. Gause[8], Martin J. Blaser[3], Richard A. Bonneau[9], Yvonne AL Lim[4,§], P'ng Loke[3,§,**], and Ken Cadwell[1,3,§,**]

Correspondence to: Ken.Cadwell@med.nyu.edu

Png.Loke@nyumc.org

limailian@um.edu.my

This PDF file includes:

Materials and Methods with references

Figs. S1 to S11

Table S1 to S4

Materials and Methods

Mice. *Nod2*$^{-/-}$, *Rag1*$^{-/-}$, *Nod2*$^{-/-}$*Rag1*$^{-/-}$, and *Stat6*$^{-/-}$ mice on the C57BL/6 background were previously described and bred onsite in an MNV/Helicobacter-free specific pathogen free (SPF) facility at NYU School of Medicine (*6, 23*). Wild-type (WT) C57BL/6 mice were purchased from Jackson Laboratory and bred onsite to generate controls for experiments. Bone marrow chimeras were generated by lethally irradiating 8-week old female recipient mice (1100 CGy in two divided doses) followed by i.v. injection of $5x10^6$ T cell-depleted bone marrow cells from donor female mice. For anti-CD4 treatment, mice were infected with *T. muris* and injected IP with 0.25mg anti-CD4 (clone GK1.5) or anti- IgG2b isotype control (clone LTF-2) (BioXCell) every 3 days for 21 days and animals were sacrificed on day 21. For anti-IL13 treatment, mice were infected with *H. polygyrus* and injected IP with 0.15mg anti-IL13 (clone 1316H) (eBioscience) or anti-IgG1 isotype control (clone 43414) (R&D systems) starting on day 3 post- infection every 3 days for 12 days and animals were sacrificed on day 12. For recombinant IL- 13 and IL-4 experiments, mice were injected I.P. with a complex of 25μg of anti-IL13 (eBio1316H) and 5μg of recombinant IL-13 (Peprotech cat#210-13), or a complex of 25μg of anti-IL4 (BioXcel) and 5μg of recombinant IL-4 (Peprotech cat#214-14) for 21 days and mice were sacrificed on day 21. For co-housing experiments, 2 mice treated with *T. muris* were placed in the same cage with 3 uninfected mice for 35 days. All animal studies were performed according to protocols approved by the NYU School of Medicine Institutional Animal Care and Use Committee (IACUC).

Human studies. The human study was approved and granted these ethical considerations (i.e., MEC Ref. No. 824.11 and No. 943.14) by the Ethics Committee of the University Malaya Medical Centre (UMMC), Malaysia, prior to the commencement of the study. The fieldwork was approved by the Department of Orang Asli Development (JAKOA) and prior permission was obtained from the Tok Batin (chieftain) of the Kuala Pangsun village before the study was conducted in the village. Kuala Pangsun village (101.88oE longitude, 3.21oN latitude) is situated

in Hulu Langat district, the fifth largest district in Selangor state, Malaysia. The selection of this village was favorable in terms of logistics and feasibility, coupled with good cooperation from the villagers. In order to assess the effect of helminths on microbial communities, a pre-post study was conducted among the 53 subjects in the Kuala Pangsun village. The villagers were requested to be involved in the pre-treatment study in May 2013, which involved the examination of intestinal helminth infections in these villagers before anthelminthic treatment was administered to them. One week after the pre-treatment screening, triple oral dose of albendazole (ABZ) tablets (3x400mg, given over 3 consecutive days), produced by GlaxoSmithKline (London, UK), was distributed to all participants. The post-treatment fecal samples were collected 20 days after treatment administration. A 21-day time point was selected to examine the short-term effects of intestinal helminth expulsion after anthelminthic treatment on gut microbiota.

Prior to sample collection, their consent was obtained in written form either through signature or thumbprint. Then, the participants were given screw-capped containers labeled with names. Containers with participants' samples were assembled back the following day and immediately frozen on dry ice. Fecal samples were then transported using dry ice to the Department of Parasitology, Faculty of Medicine, University of Malaya. Then, fecal samples collected in screw-capped containers were separated into two portions: a) preserved in 2.5% potassium dichromate and stored at 4°C (for intestinal helminth screening), b) stored in 1.5ml microcentrifuge tube and kept in -80°C prior to DNA extraction steps (for microbiome analysis). Formalin-ether sedimentation technique was used to determine the presence and absence of the intestinal helminth as previously described (*24*). The intensity of infection was determined by Kato Katz technique and results were recorded as egg per gram (epg), using the cut off determined by the WHO (*3*). Then, the fecal smears were observed using a light microscope under the magnification of 100X and 400X and the cure rate (percentage of helminth infected individuals who became egg-negative after albendazole treatment) was determined. In addition, 19 stool samples were

collected randomly from University of Malaya as urban controls for this study with 6 Chinese, 1 Sabahan, 1 Sarawakian, 7 Indians, 2 Malays, and 2 Yemeni. Although of different races/nationality, these participants have been residing in Kuala Lumpur for over 5 years, and were exposed to a similar variety of food.

Microscopy and flow cytometry. Intestinal sections were prepared and stained with PAS/Alcian blue as previously described (*1*). The number of granule-containing goblet cells in the villi were quantified by counting positively-stained cells displaying the characteristic goblet morphology (*1*). Immunohistochemistry analysis of pSTAT6 was performed on formalin fixed paraffin-embedded small intestinal sections and the number of pSTAT6 cells in each frame under 20X magnification were quantified. Immuno-fluorescence analysis of Reg3β was performed by staining formalin fixed paraffin-embedded small intestinal sections, which were imaged using the Zeiss Axioplan epifluorescence microscope and quantified using ImageJ as previously described (*1*). Intra-epithelial and lamina propria lymphocytes were isolated from the small intestine and colon as previously described (*1*). Lymphocytes were stimulated for 4 hours with a cell stimulation cocktail of PMA, ionomycin, brefeldin A and monensin from eBioscience. Stimulated cells were stained with anti-CD3ε PerCP, anti-TCRβ PE-Cy5, anti-CD8α PE-Cy7, anti CD4 APC-Cy7, anti- IFNγ APC, anti-IL17 PE, anti-IL4 APC and their respective isotype controls from Biolegend, and, anti-IL-13 FITC, anti-IL-10 PE and their respective isotype controls from eBioscience. Fixation and permeabilization buffers from Biolegend were used for intracellular cytokine staining, and a fixable live/dead stain from Biolegend was used to exclude dead cells. For nuclear staining, unstimulated cells were stained with anti-CD45 PerCP, anti-TCRβ BV510, anti CD4 APC-Cy7, anti-Ki67 Alexa700, anti-CD25 APC, anti-Neuropilin1 BV421, anti-Helios PE-Cy7, anti-RORγt PE, anti-Foxp3 FITC, and their respective isotype controls from Biolegend using the Foxp3 staining kit (eBioscience). Flow cytometric analysis was performed on an LSR II (BD biosciences) and analyzed using FlowJo software from TreeStar.

Bacteria. Fecal and tissue-associated *B. vulgatus* were quantified by dilution plating on selective BBE agar (BD) for 24-48 hours in an anaerobic chamber (AS-580, Anaerobe Systems). To quantify tissue-associated bacteria, one cm of small intestinal (ileum) tissue was flushed with PBS, cut open longitudinally, washed in PBS and homogenized. For inoculation into mice, *B. vulgatus* (*1*) and *L. johnsonii (16)* were anaerobically cultured at 37ºC in PYG broth for 48 hrs and Lactobacilli MRS broth for 24 hrs, respectively. The mix of 17 human Clostridia species previously described (*15*), were anaerobically cultured at 37ºC in PYG broth for 48 – 96 hours individually and combined prior to inoculation. Mice were orally gavaged with a 100µl broth solution containing 1 x 10^8 cfu Clostridia mixture, 1 x 10^8 cfu *L. johnsonii*, or sterile broth every 3 days for 3 weeks. For *in vitro* mucus experiments, partially purified mucin from porcine stomach (M1778, SigmaAldrich), predominantly consisting of muc2, was dissolved in a solution of 0.1M sodium acetate (pH 5) at a concentration of 20mg/ml. Equal amounts of Clostridiales species (*Clostridium (#28), Ruminococcus (#13), and Erysipelatoclostridium (#18)*) or *B. vulgatus* were added to PYG broth with increasing concentrations of mucin (50, 100, 250 or 500µg/ml) or vehicle control (0.1M sodium acetate) and plated at 1, 3, 6, and 9 hours post addition of mucin.

Helminth infection. Maintenance of the *T. muris* lifecycle was carried out as described previously (*25*). Mice were infected with ~25 embryonated eggs by oral gavage and sacrificed at D35 post infection. Worm burdens were assessed as described previously (*26*). In all conditions analyzed, successful chronic infection was established and equal worm burdens were confirmed. For *Heligmosomoides polygyrus* infection (*27*), mice were infected with approximately 200 L3 larvae via oral gavage and sacrificed at D12 (acute) or D35 (chronic) post infection.

Piroxicam treatment. *T. muris* infected and uninfected WT and *Nod2*$^{-/-}$ mice were treated with 60mg/250g and 80mg/250g of piroxicam as previously described (*1*). H&E stained small intestinal sections of mice treated with piroxicam were used for histopathological scoring in a

blinded fashion. Each mouse was given an individual cumulative score based on the following criteria: number of focal ulcers (0 = none, 1 = 1, 2 = 2 and so on), number of abscesses (0 = none, 1 = 1, 2 = 2 and so on), the extent of epithelial hyperplasia (0 = none, 1 = elongated villi and crypts, 2 = severe hyperplasia where the crypt villus axis is 2 times higher than the crypt villus axis in untreated mice), the presence of immune infiltrates (0 = none, 1 = pericryptal infiltrates, 2 = submucosal infiltrates), and villus blunting (0 = none, 1-2 = moderate blunting, 3-4 = severe blunting). The presence of macroscopic abnormalities such intestinal bleeding, and/or intestinal perforation in each mouse was also tabulated. For bacterial translocation assay, one cm of flushed small intestinal ileum, the entire spleen, and MLNs from piroxicam treated mice were homogenized and plated on blood agar plates and incubated for 24 hours at 37ºC.

Gene expression analysis. RNA was isolated from one cm of small intestinal tissue (ileum) from *Nod2*$^{-/-}$ mice treated with rIL-13 or PBS, or WT mice treated with PBS for normalization. RNA library was prepared using the TruSeq stranded total RNA, with RiboZero Gold kit and samples were sequenced on the Illumina HiSeq to generate single-end 50bp reads. Raw sequencing reads were aligned to the mouse reference genome mm10 and the RefSeq reference transcriptome using TopHat (v2.0.12) with the alignment parameter --library-type fr-firststrand (*28*). Reads with mappability score (MAPQ) <30 were removed. The total number of filtered reads were counted for each gene using htseq-count with the parameters --stranded=reverse and --mode=union. Singleton genes (i.e. genes with total count < 1) were filtered and the resulting count matrix was used for differential analysis using the DESeq2 workflow with default parameters(*29, 30*). Differential genes with rIL-13 treatment were identified at FDR 10%. Pathway analysis was perfomed using the DAVID platform as previously described(*6*). This data has been made publicly available and can be accessed using the GEO accession number GSE76504.

16S library preparation. DNA was isolated from stool samples using the NucleospinSoil Kit (Macherey-Nagel). Bacterial 16S rRNA gene was amplified at the V4 region using a modified

protocol from Caporaso et al (*31*). The forward primer construct contained the 5' Illumina adapter, the forward primer pad, a two-base linker ('GT') and the 515F primer (5'- AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT ATG GTA ATT GTG TGC CAG CMG CCG CGG TAA -3'). The reverse primer construct contained the 3' Illumina adapter, a unique 12-base error-correcting Golay barcode, the reverse primer pad, a two-base linker sequence ('CC') and the 806R primer (5'- CAA GCA GAA GAC GGC ATA CGA GAT NNN NNN NNN NNN AGT CAG TCA GCC GGA CTA CHV GGG TWT CTA AT -3'). Cycling protocol consisted of 94°C for 3min, 35 cycles of 94°C for 45s, 50°C for 60s and 72°C for 90s, with a final extension of 72°C for 10min. Amplification was performed in triplicates and the pooled amplicon was purified with QIAgen PCR purification kit. The purified amplicon library was pooled at equimolar ratio and sequenced on the Illumina MiSeq with a 2 x 150 cycle run (Illumina, San Diego CA, USA).

16S sequences analyses. Sequencing read mates of the 16S library were joined using the fastq-join function from EA-utils (*32*). The joined reads were processed using the Quantitative Insights Into Microbial Ecology (QIIME) software package (*33*). The split_library.py function was first used for demultiplexing, in addition to performing quality filter with default parameters (minimum quality score of 25, minimum/maximum length of 200/1000, no ambiguous bases allowed and no mismatches allowed in the primer sequence). Operational Taxonomic Units (OTUs) were defined using a combination of closed reference and *de novo* sequence clustering methods (pick_open_reference_otus.py workflow in QIIME). The Greengenes reference collection (version 13_5) was used as reference sequences and similarity threshold was defined at 97%. The resulting OTU table was filtered for singletons before downstream analyses. Alpha diversity analysis was done using the metrics observed OTUs, Shannon index and chao1 (*34, 35*). The QIIME alpha rarefaction workflow (alpha_rarefaction.py) was used with default parameters on an OTU table that was first rarefied to the minimum sampling depth. Beta diversity was calculated using unweighted UniFrac distance performed on an uneven OTU table (*36, 37*).

Principle Coordinate Analyses (PCoA) was performed on the UniFrac distance matrix and the resulting PCoA plot visualized using the Emperor graphics program (*38*). The LDA Effect Size (LEfSe) algorithm (http://huttenhower.sph.harvard.edu/galaxy/) was used to identify differentially abundant taxa in different biological groups at a threshold LDA score described in the legends (*39*).

Statistical analysis. An unpaired two-tailed t test was used to evaluate differences between two groups. A paired two-tailed t test was used to evaluate differences between different time points in the same group. An ANOVA was used to evaluate experiments involving multiple groups with the Holm-Sidak multiple comparisons test. For contingency tables, Fisher's exact test was used. For experiments requiring non-parametric analyses, the Wilcoxon-Mann-Whitney test was used. For weight loss analysis, an ANOVA with the Holm-Sidak multiple comparisons test was used to evaluate the data obtained from analyzing the area under the curve for each individual mouse.

Bioinformatic analysis

Sparse and compositionally-robust PLS regression

We sought to detect associations between specific taxa in fecal microbiota communities and host-side measures, while reducing the detection of statistically spurious associations. This pipeline: was (i), the compositionally-robust centered log-ratio transformation (clr) (*40*) of OTU relative abundance data (with a single pseudocount added prior to normalization) and (ii), estimation of a sparse linear model via Partial Least Squares (PLS) regression, to model high-dimensional and multi-collinear feature/responses (e.g. OTUs, taxa, and host covariates). We dummy coded categorical variables into values of (-1 or 1) to indicate if a sample belongs to that class.

We then applied *L*1-penalized partial least squares regression (*41-43*) to fit a bi-linear model. The number of latent components in the sPLS model are fixed to the number of non-zero singular

values in the cross-covariance matrix. Model sparsity is controlled via the scalar parameter η that weights the influence of the *L*1 penalty. We used a two-stage approach to find a sparse set of significant OTU-phenotype associations. In the first stage, we used stability selection approach to regularization selection (StARS (*44*)) to determine the sparsity parameter η; StARS has been previously shown to be competitive for graphical model problems of similar complexity and scale (*44*). We rebuilt the sPLS model over 100 random subsets of the data over a range of values for η, calculating the fraction of data subsets that included a given OTU in the support (i.e., the non-zero model coefficients) at each η. We then computed a summary statistic of overall model variability to select the most stable model that exceeds the variability threshold (0.1%) (*44*). In the second stage, we assessed the statistical significance of individual OTUs in the model by computing empirical p-values over 30,000 bootstrapped PLS models on the StARS-selected support. We compared the models to an empirical null model (generated by fitting randomized permutations of the data), which yields a p-value for each OTU-host phenotype pair.

For the experiments that relied on repeated measures design (i.e. Trichuris deworming), we incorporated an additional step into the pipeline. We decomposed the clr-transformed OTU compositions as well as host responses data into the relevant 'within-subject' components using a one-factor variance decomposition (*45*). The within-subject component captures experimental perturbation effects by subtracting between-subject variances, and is directly proportional to the change in data levels over the single repeated measure. We applied the sPLS model directly to the within-subject variances as described (*45*) and implemented in the mixOmics package in R (*46*).

We used routines from the spls and caret libraries in R and developed a custom package (which includes methods for the full pipeline, a similar approach for discriminant analysis (*17*) and biplots) called compPLS [software and additional methods are available at http://github.com/zdk123/compPLS].

This approach identifies a small set of bacterial OTUs for which a linear combination of abundance changes accurately models concurrent changes in Trichuris trichuria egg burden (Δ

Trichuris). Though this is a multivariate model, we report individual OTU - Δ Trichuris relationships by filtering empirical p-values of the model coefficient ($\alpha=10^{-2}$) and the direction of association from sign of the coefficient (Table S4).

We visualized these results in several ways. For continuous responses, such as Δ Trichuris, we report within-sample prediction error (r^2) in a residual plot. For categorical responses (e.g. Sex) the model predicts a continuous value, fit to the dummy coded value. To show within-model error for classification, we used a varying threshold of these values as a cutoff to predict membership in this class and report the True Positive vs False Positive rate (Receiver Operator Characteristic - ROC - curve) as this threshold is varied. Useful models show better than random classification (AUC=.05).

Finally, to visualize pairwise relationships between taxa and ΔTrichuris, we show data fit to the model (i.e. decomposed within-subject variances, scaled and centered) as scatterplots. We show model fits as overlaid straight lines, slopes are fitted model coefficients, to show the relative contribution of the OTU to the overall prediction.

We visualized multivariate relationships between OTUs and ΔTrichuris learned from the sPLS model using biplots. The subspace where OTUs and response maximally covaries is learned by the PLS model, the data is projected onto the loadings (sparse set of OTUs), which are the model 'scores'. Loadings are visualized as vectors, representing each OTU, and colored at the order level, length of each vector is proportional to the contributing variance of that OUT and the angle between each loading vector indicates the correlation between those OTUs in the respective space (0° - perfect correlation, 90° - no correlation, 180° perfect anti-correlation). Each PLS component represents the linear combination of OTUs contributing to the OTU- ΔTrichuris covariation in that space, and since each successive PLS component is learned from the residuals of the data projected onto the previously-learned component, PLS components are orthogonal.

Estimation of Microbial association networks

We selected several publicly available datasets to compare to the Orang Asli (OA) (n=45 before and after deworming) microbial networks. These were American Gut Project (AGP) (*22*) (fecal, n=3671), AGP samples [self-identified] IBD only (fecal, n=178), the Human Microbiome Project (HMP) data (v35 reads) (*21*) (fecal, n=402) and the RISK IBD cohort (*20*) (fecal and GI biopsies, n=913). AGP and HMP project data was obtained from biocore project [https://github.com/biocore/American-Gut] and RISK cohort microbiome data was obtained from the Qiita repository [http://qiita.microbio.me/] (Study IDs 1939 and 1998).

OTUs that appeared in fewer than 37% (chosen for consistency with previous studies of AGP (*19*)) of samples across the entire dataset were filtered out for this analysis. We then applied Sparse Inverse Covariance estimation for Ecological ASsociation Inference (SPIEC-EASI) (*19*) to examine the network model from each cluster, where nodes in the network are OTUs and edges are inferred relationships between OTUs it its environment. We ran SPIEC-EASI, using Meinshausen-Bühlmann (MB) neighborhood selection, and network model selection via StARS, using a variability threshold of 0.05%.

Over the set of OTUs retained (p) the number of edges inferred (e) for each dataset was: OA-pre, p=422 and e=2143, OA-post, p=424 and e=2048; AGP, p=243 and e=781; AGP-IBD, p=152 and e=327; HMP, p=108 and e=141; RISK, p=246 and e=890.

To characterize relationships at the order level, we summarized both positive and negative model coefficients (symmetrizing MB edges by taking the maximum absolute value of coefficient) between OTU pairs. We then report the fraction of positive edges between all OTUs of the same order in a network diagram, plotting edge color and the majority sign (green – positive, red – negative). Node size is proportional to total number of edges for that order (therefore reflecting the number of taxa as well as overall connectivity).

Classification of Orang Asli into responders and non-responders to treatment

We classified the response to treatment among Orang Asli samples by measuring the difference between log-transformed Trichuris levels (with a single pseudocount added to counts) before and after treatment. For patient samples that had parasites prior to deworming, we partitioned change in Trichuris burden around randomly selected cluster mediods (pam) and found that response to deworming clustered into two distinct groups based on a threshold of -Δ0.03 (above which we called non-responders and the rest being classified as non-responders). Using after treatment samples, we fit an sPLS model to clr-transformed compositions to predict log-transformed Trichuris trichuria worm burden. We compared the within-sample fit among responders (r^2=0.693) to the out-of-sample fit of Non-responders (r^2=0.01).

References:

23. U. M. Gundra *et al.*, Alternatively activated macrophages derived from monocytes and tissue macrophages are phenotypically and functionally distinct. *Blood* **123**, e110-122 (2014).
24. WHO, Guidelines for the evaluation of soil-transmitted helminthiasis and schistosomiasis at community level. A guide for managers of control programmes. . *Geneva: World Health Organization*, (1998).
25. D. Wakelin, Acquired immunity to Trichuris muris in the albino laboratory mouse. *Parasitology* **57**, 515-524 (1967).
26. K. J. Else, D. Wakelin, D. L. Wassom, K. M. Hauda, MHC-restricted antibody responses to Trichuris muris excretory/secretory (E/S) antigen. *Parasite Immunol* **12**, 509-527 (1990).
27. R. M. Anthony *et al.*, Memory T(H)2 cells induce alternatively activated macrophages to mediate protection against nematode parasites. *Nat Med* **12**, 955-960 (2006).
28. D. Kim *et al.*, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36 (2013).
29. S. Anders, P. T. Pyl, W. Huber, HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).
30. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
31. J. G. Caporaso *et al.*, Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A* **108 Suppl 1**, 4516-4522 (2011).
32. E. Aronesty. (2011).
33. J. Caporaso *et al.*, QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**, 335 - 336 (2010).
34. C. E. Shannon, A mathematical theory of communication. *Bell System Technical Journal* **27**, 379-423, 623-656 (1948).
35. A. Chao, Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of statistics*, 265-270 (1984).

36. J. Chen, K. Bittinger, E. Charlson, C. Hoffmann…, Associating microbiome composition with environmental covariates using generalized UniFrac distances. ..., (2012).
37. C. Lozupone, M. E. Lladser, D. Knights, J. Stombaugh, R. Knight, UniFrac: an effective distance metric for microbial community comparison. *The ISME journal* **5**, 169-172 (2011).
38. Y. Vazquez-Baeza, M. Pirrung, A. Gonzalez, R. Knight, EMPeror: a tool for visualizing high-throughput microbial community data. *GigaScience* **2**, 16 (2013).
39. N. Segata *et al.*, Metagenomic biomarker discovery and explanation. *Genome Biol* **12**, R60 (2011).
40. J. Aitchison, The statistical analysis of compositional data. *Chapman and Hall, London* **New York**, (1986).
41. D. Chung, S. Keles, Sparse partial least squares classification for high dimensional data. *Stat Appl Genet Mol Biol* **9**, Article17 (2010).
42. H. Chun, S. Keles, Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *J R Stat Soc Series B Stat Methodol* **72**, 3-25 (2010).
43. D. R. Kim-Anh Lê Cao, Christèle Robert-Granié, and Philippe Besse., A sparse PLS for variable selection when integrating omics data. *Statistical applications in genetics and molecular biology,* **7**, (2008).
44. H. Liu, K. Roeder, L. Wasserman, Stability Approach to Regularization Selection (StARS) for High Dimensional Graphical Models. *Adv Neural Inf Process Syst* **24**, 1432-1440 (2010).
45. B. Liquet, K. A. Le Cao, H. Hocini, R. Thiebaut, A novel approach for biomarker selection and the integration of repeated measures experiments from two assays. *BMC Bioinformatics* **13**, 325 (2012).
46. I. G. Kim-Anh Le Cao, Sebastien Dejean with key contributors Florian Rohart, Benoit Gautier, contributions from Pierre Monget, Jeff Coquery, FangZou Yao, and Benoit Liquet, mixOmics: Omics Data Integration Project. (2015).
47. R. K. Grencis, Immunity to helminths: resistance, regulation, and susceptibility to gastrointestinal nematodes. *Annu Rev Immunol* **33**, 201-225 (2015).

Supplementary figure 1

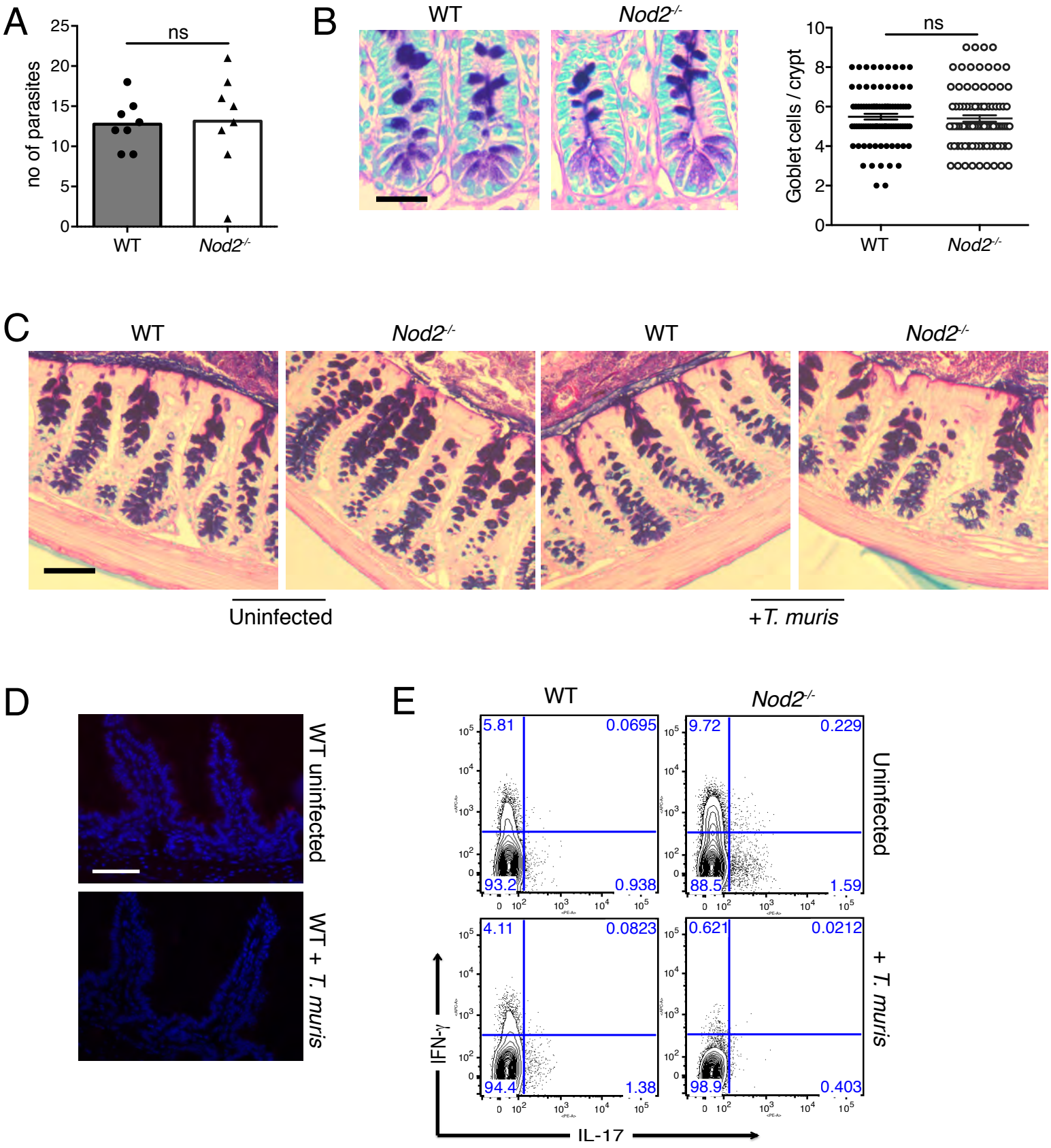
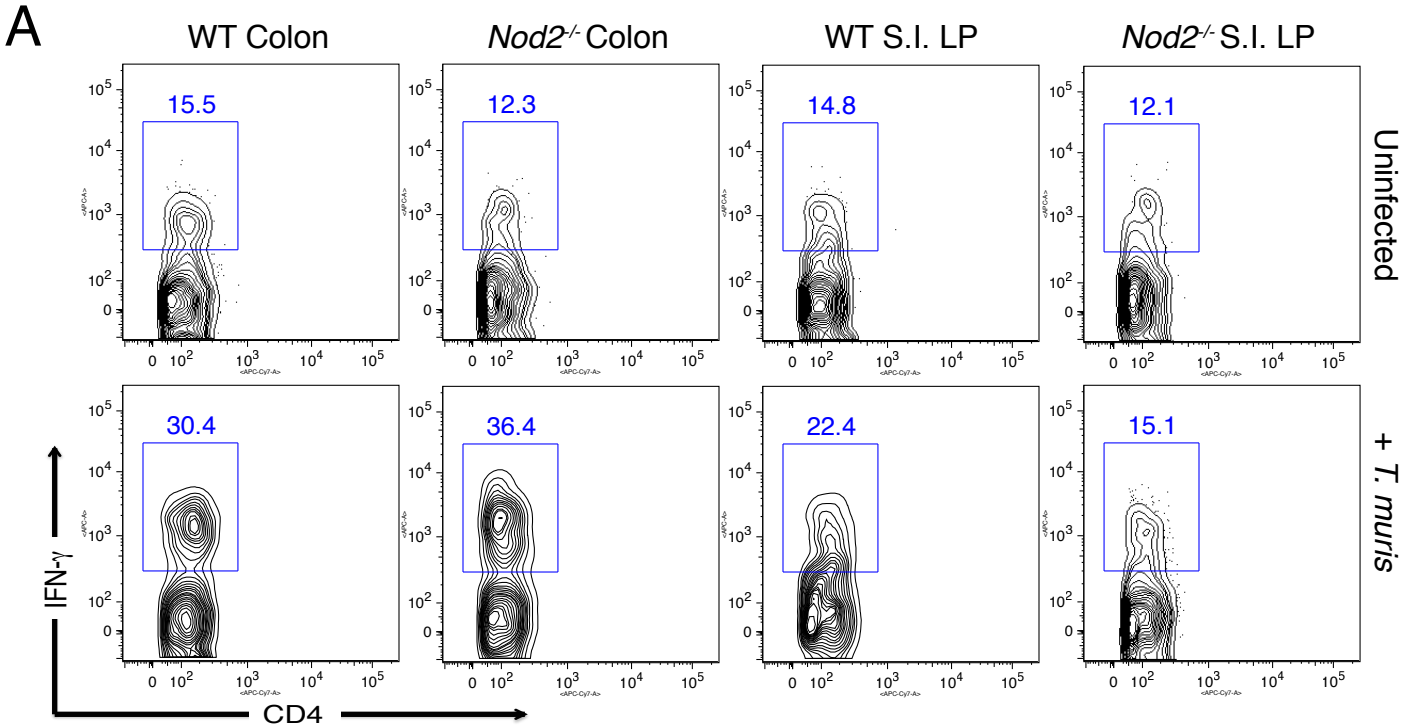
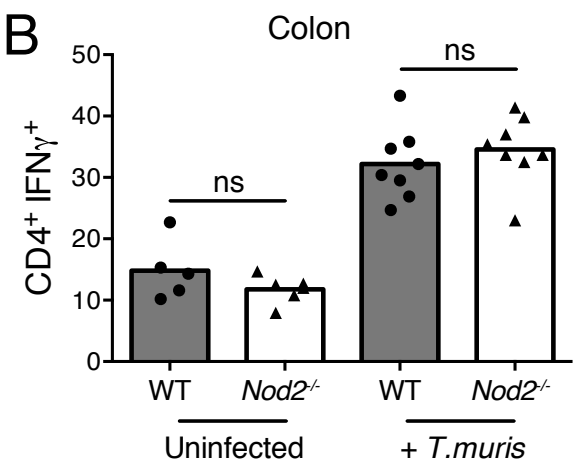


Figure S1: *T. muris* can reverse intestinal abnormalities in *Nod2*$^{-/-}$ mice.

(A) Quantification of *T. muris* in the cecum and large intestine of WT and *Nod2*$^{-/-}$ mice 35 days post infection demonstrates successful colonization of both genotypes (n=8 mice per genotype). **(B)** Representative images and quantification of goblet cells in PAS/Alcian blue stained sections of small intestinal crypts in uninfected WT and *Nod2*$^{-/-}$ mice. We have previously described that goblet cell defects in the villi of *Nod2*$^{-/-}$ small intestine is due to increased IFN-γ production by IELs (*6*). In contrast to the villi, small intestinal crypts display no differences in goblet cell numbers, likely due to the scarcity of IELs. Scale bar represents 50μm. **(C)** Representative images of PAS/Alcian blue stained sections of the colon demonstrating similar goblet cell numbers in uninfected and *T. muris* infected WT and *Nod2*$^{-/-}$ mice. Scale bar represents 100μm. **(D)** Immunofluorescence (IF) analysis of Reg3β in uninfected and *T. muris* infected WT small intestinal sections. Samples from *Nod2*$^{-/-}$ mice from the same experiment are in Figure 1C with the quantification of all four conditions in Figure 1D. Scale bar represents 100μm (n≥8 mice per genotype). **(E)** Representative flow cytometry plot of CD8^{+} IELs in uninfected and *T. muris* infected WT and *Nod2*$^{-/-}$ mice stained for intracellular IFN-γ and IL-17 expression after stimulation with PMA and ionomycin gated on CD3^{+} live cells (n≥11 mice per genotype). Data are representative of at least two independent experiments. Each data point represents an individual mouse, bar denotes mean in (A). Data are represented as mean ± SEM for individual colon crypts in (B).

Supplementary figure 2





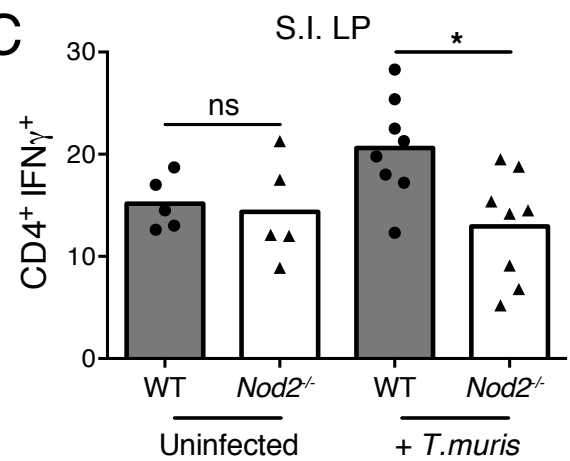


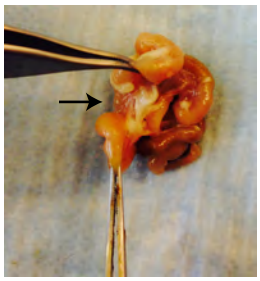
Figure S2: Chronic *T. muris* infection induces an increase in IFN-γ producing CD4$^+$ T cells in the intestinal lamina propria.

(A-C) Representative flow cytometry plot (A) and quantification of CD4$^+$ T cells in the colon (B) and small intestinal lamina propria (C) of uninfected and *T. muris* infected WT and *Nod2*$^{-/-}$ mice stained for intracellular IFN-γ expression after stimulation with PMA and ionomycin (n≥5 mice per genotype). CD4$^+$ T cells in the colon of infected WT and *Nod2*$^{-/-}$ mice, displayed an increase in IFN-γ production, an observation that is consistent with the effect of chronic *T. muris* infection of mice on the C57BL/6 background (*47*). However, in the small intestinal lamina propria, only CD4$^+$ T cells in WT mice display an increase in IFN-γ production post *T. muris* infection. *$p<0.05$ by ANOVA with Holm-Sidak multiple comparisons test for (C). Data are representative of at least two independent experiments. Each data point represents an individual mouse, bar denotes mean.

Supplementary figure 3

A

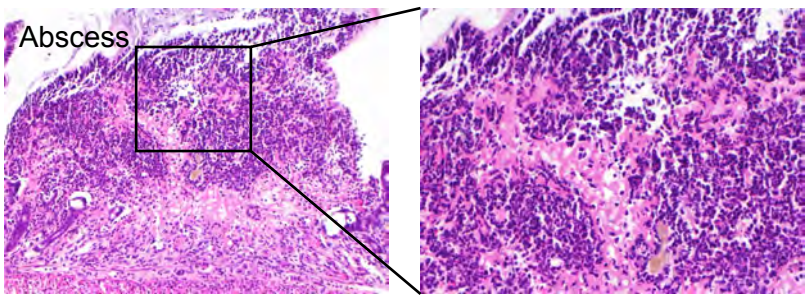
B



C

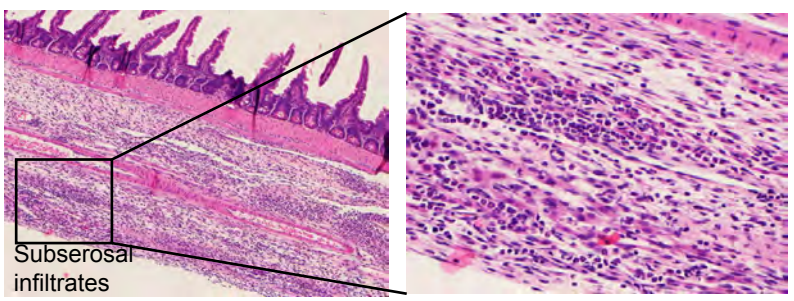
	Piroxicam only		+*T. muris*	
	WT	*Nod2*$^{-/-}$	WT	*Nod2*$^{-/-}$
Intestinal bleeding	0/10 0%	6/10 ** 50%	0/10 0%	0/12 0%
Attachment to other tissues	0/10 0%	7/10 *** 75%	0/10 0%	0/12 0%

D

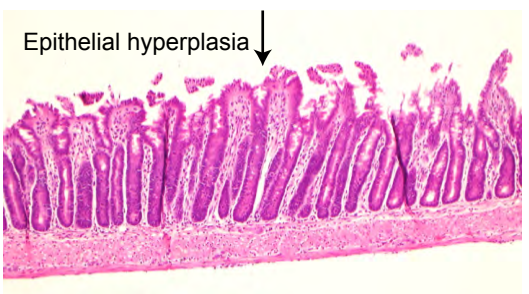


E

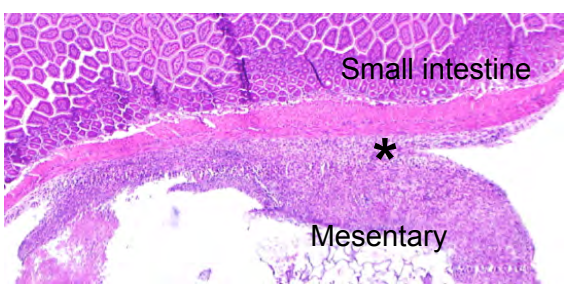
F



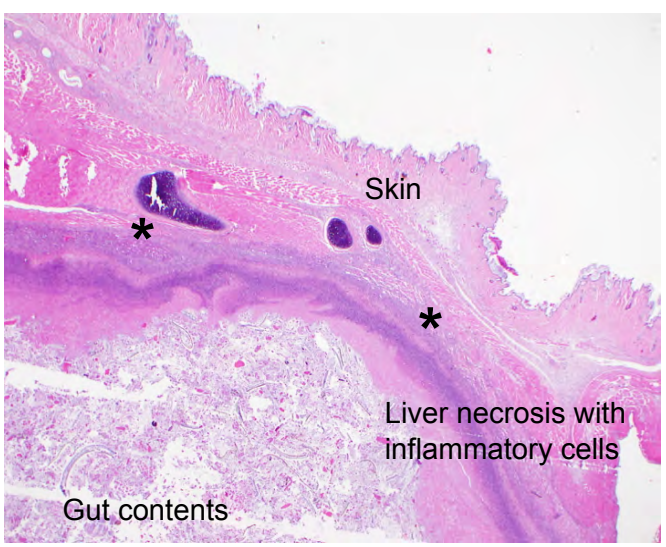
G



H



I



J

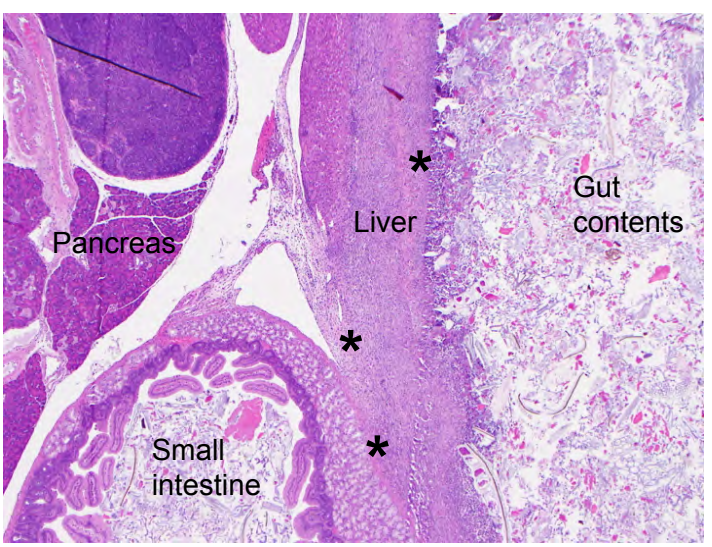


Figure S3: *T. muris* protects *Nod2*$^{-/-}$ mice from piroxicam-induced pathologies.

(A) Representative image comparing the small intestine of uninfected (right) or *T. muris* infected (left) *Nod2*$^{-/-}$ mice treated with piroxicam. Black arrows denote the presence of live parasites in the cecum of *T. muris* infected *Nod2*$^{-/-}$ mice. In contrast to the *T. muris* infected *Nod2*$^{-/-}$ or WT mice, or uninfected WT mice, the small intestine from uninfected *Nod2*$^{-/-}$ mice cannot be straightened without tearing due to fusion of the organ with itself and surrounding tissue within the peritoneal cavity including mesenteric fat and cecum. **(B)** Representative image of fusion of the small intestinal wall with surrounding tissue (denoted by black arrow) in uninfected *Nod2*$^{-/-}$ mice treated with piroxicam. **(C)** Table representing quantification of small intestinal bleeding and attachment of the intestine to the surrounding tissue of the peritoneal cavity as in (A) and (B) in uninfected and *T. muris* infected WT and *Nod2*$^{-/-}$ mice treated with piroxicam. **(D-J)** Representative H&E-stained sections of the small intestine demonstrating the presence of the following pathologies detected at higher frequency in *Nod2*$^{-/-}$ mice treated with piroxicam: abscesses (higher magnification images show presence of infiltrating neutrophils and plasma cells, fibrin deposition, and cell debris, characteristic of abscesses) (D), subserosal infiltrates (higher magnification image shows presence of neutrophils and plasma cells infiltrating the subserosa) (E), villus blunting (F), and epithelial hyperplasia (G). Occurrence of these pathologies was used to generate the pathology score in Figure 1G as described in the Methods section. **(H-J)** Representative images of H&E-stained sections prepared from fused organs harvested from *Nod2*$^{-/-}$ mice treated with piroxicam as in (B) demonstrate intestinal perforation. Mesenteric tissue was observed penetrating into small intestinal tissue (asterisk denotes fusion of the two tissue types) (H). The presence of intestinal contents was observed in adjoining tissues such as the liver and skin (asterisks denote areas of fusion between the different tissue types; liver tissue surrounding the intestinal contents have increased immune infiltrates and abscesses) (I) and (J). **$p<0.01$, ***$p<0.001$ by Fisher's exact test for (C). Image magnification at 2x, 10x, and 40x

in (D); 5x, and 40x in (E); 5x in (F), (G), and 2x in (H)-(J). Data are representative of at least two independent experiments.

Supplementary figure 4

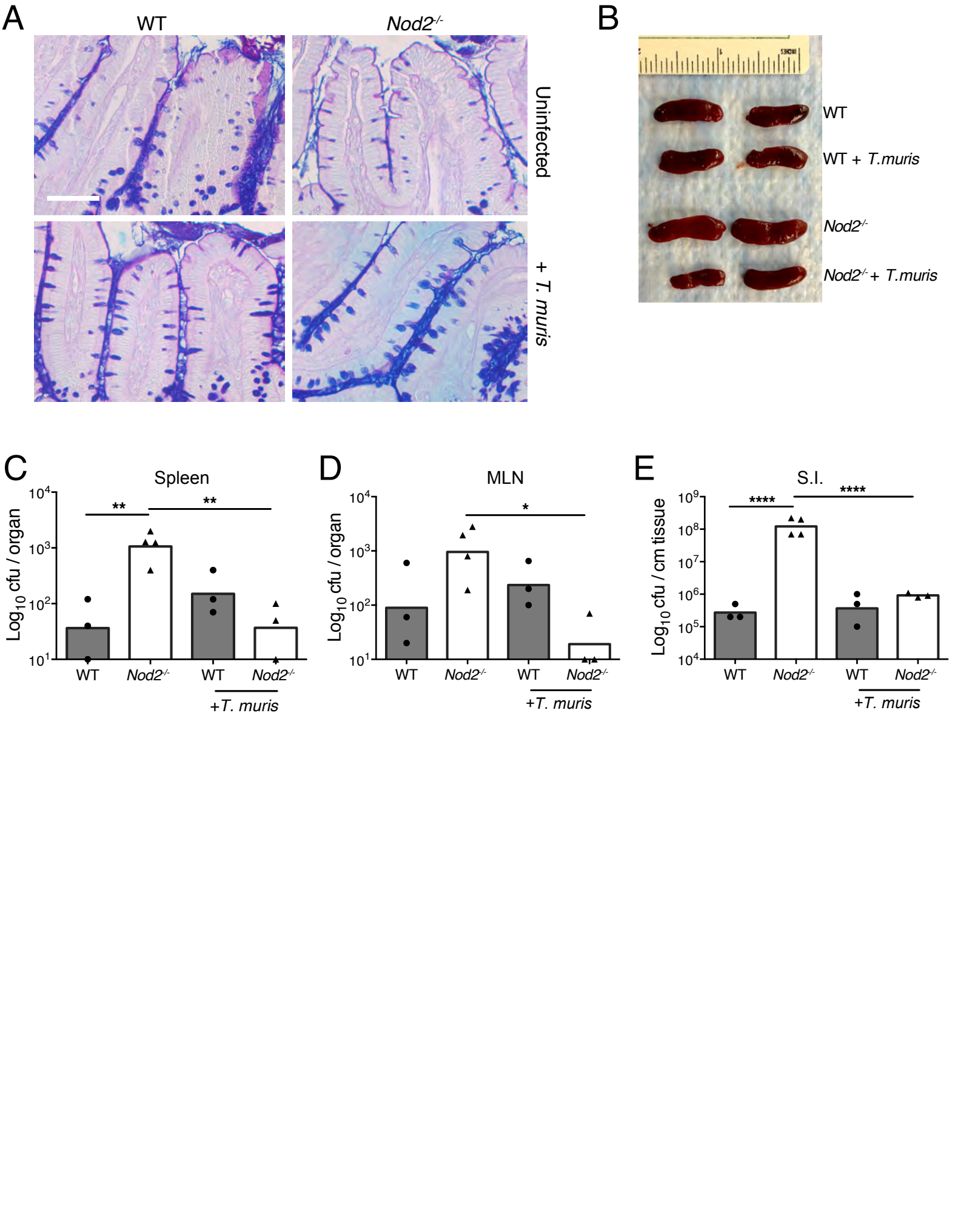


Figure S4: *T. muris* protects piroxicam-treated *Nod2*$^{-/-}$ mice from disease manifestations. **(A)** Representative PAS-Alcian blue-stained sections of the small intestine form uninfected and *T. muris* infected WT and *Nod2*$^{-/-}$ mice treated with piroxicam. These images show that uninfected *Nod2*$^{-/-}$ mice display a decreased presence of mucus (blue staining), which is restored upon *T. muris* infection, an observation that is consistent with a reduction in bacterial translocation following *T. muris* infection. Scale bar represents 50μm. **(B)** Spleens from *Nod2*$^{-/-}$ mice treated with piroxicam were larger compared with those from similarly treated uninfected WT and *T. muris* infected WT and *Nod2*$^{-/-}$ mice. **(C-E)** Quantification of bacterial colony forming units (cfu) in spleen (C), mesenteric lymph nodes (D), and small intestinal tissue (see methods) (E), in uninfected and *T. muris* infected WT and *Nod2*$^{-/-}$ mice treated with piroxicam ($n \geq 3$ mice per genotype). *$p<0.05$, **$p<0.01$, and ****$p<0.0001$ by ANOVA with Holm-Sidak multiple comparisons test for (C-E). Data are representative of at least two independent experiments. Each data point represents an individual mouse, bar denotes mean.

Supplementary figure 5

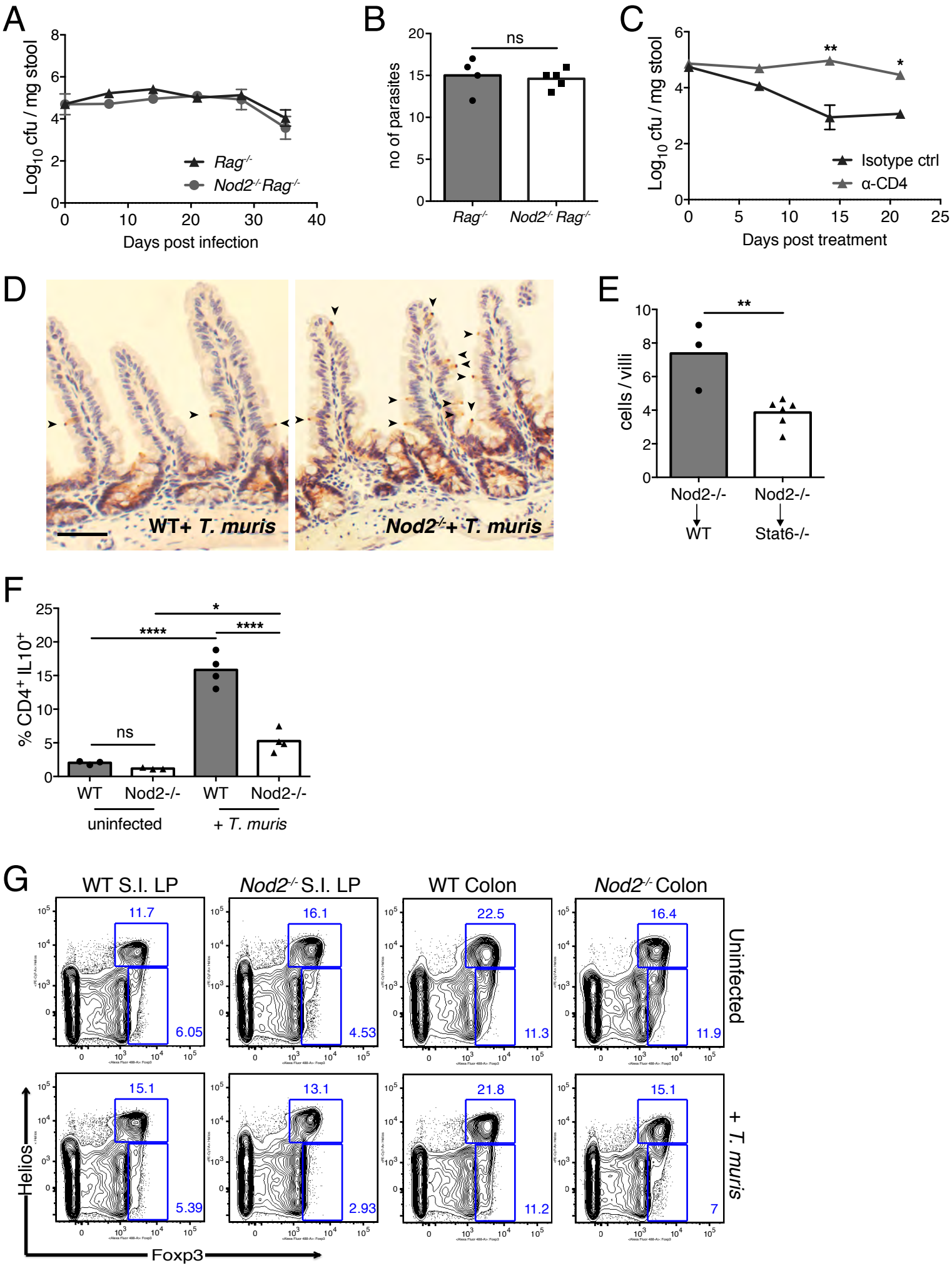


Figure S5: *T. muris* mediated protection from intestinal abnormalities is associated with CD4$^+$ T cells.

(A) Quantification of *B. vulgatus* in stool of *T. muris* infected *Rag*$^{-/-}$ and *Nod2*$^{-/-}$*Rag*$^{-/-}$ mice (n≥4 mice per genotype). **(B)** Quantification of *T. muris* in the cecum and large intestine of *Rag*$^{-/-}$ and *Nod2*$^{-/-}$*Rag*$^{-/-}$ mice from (A) demonstrating that the inability to reduce B. vulgatus burden is not due to lack of chronic parasite infection (n≥4 mice per genotype). **(C)** Quantification of *B. vulgatus* in stool of *T. muris* infected *Nod2*$^{-/-}$ mice treated with antibody to CD4 or isotype control (n≥4 mice per genotype). **(D)** Representative immunohistochemistry staining of pSTAT6 in the small intestine of uninfected and *T. muris* infected *Nod2*$^{-/-}$ mice. Quantification shown in Figure 2B. Black arrows denote pSTAT6 positive cells in the small intestinal villi. Scale bar represents 100μm. **(E)** Quantification of goblet cells displaying normal morphology in the small intestine of *T. muris* infected WT (*Nod2*$^{-/-}$ → WT) and *Stat6*$^{-/-}$ (*Nod2*$^{-/-}$ → *Stat6*$^{-/-}$) mice reconstituted with *Nod2*$^{-/-}$ bone marrow (BM) shows that Stat6 is necessary for helminth-mediated restoration of goblet cells (n≥3 mice per genotype). **(F)** Quantification of CD4$^+$ T cells expressing IL-10 in the small intestinal lamina propria of uninfected and *T. muris* infected WT and *Nod2*$^{-/-}$ mice after stimulation with PMA and ionomycin (n≥3 mice per genotype). While infected *Nod2*$^{-/-}$ mice display a marginal increase, *T. muris* infected WT mice display a substantial increase in IL-10 producing CD4$^+$ T cells, consistent with previous studies (*47*). **(G)** Representative flow cytometry plot of uninfected and *T. muris* infected WT and *Nod2*$^{-/-}$ mice stained for Helios and Foxp3 expression in CD4$^+$ live T cells, indicating that the number of natural or induced regulatory T cells were not significantly altered in the intestinal tissues of either infected WT or of *Nod2*$^{-/-}$ mice. *p<0.05, **p<0.01, and ****p<0.0001 by unpaired t-test in (C) and (E), and ANOVA with Holm-Sidak multiple comparisons test for (F). Data are representative of at least two independent experiments. Each data point represents an individual mouse, bar denotes mean in (B), (E), and (F). Data are represented as mean ± SEM in (A) and (C).

Supplementary figure 6

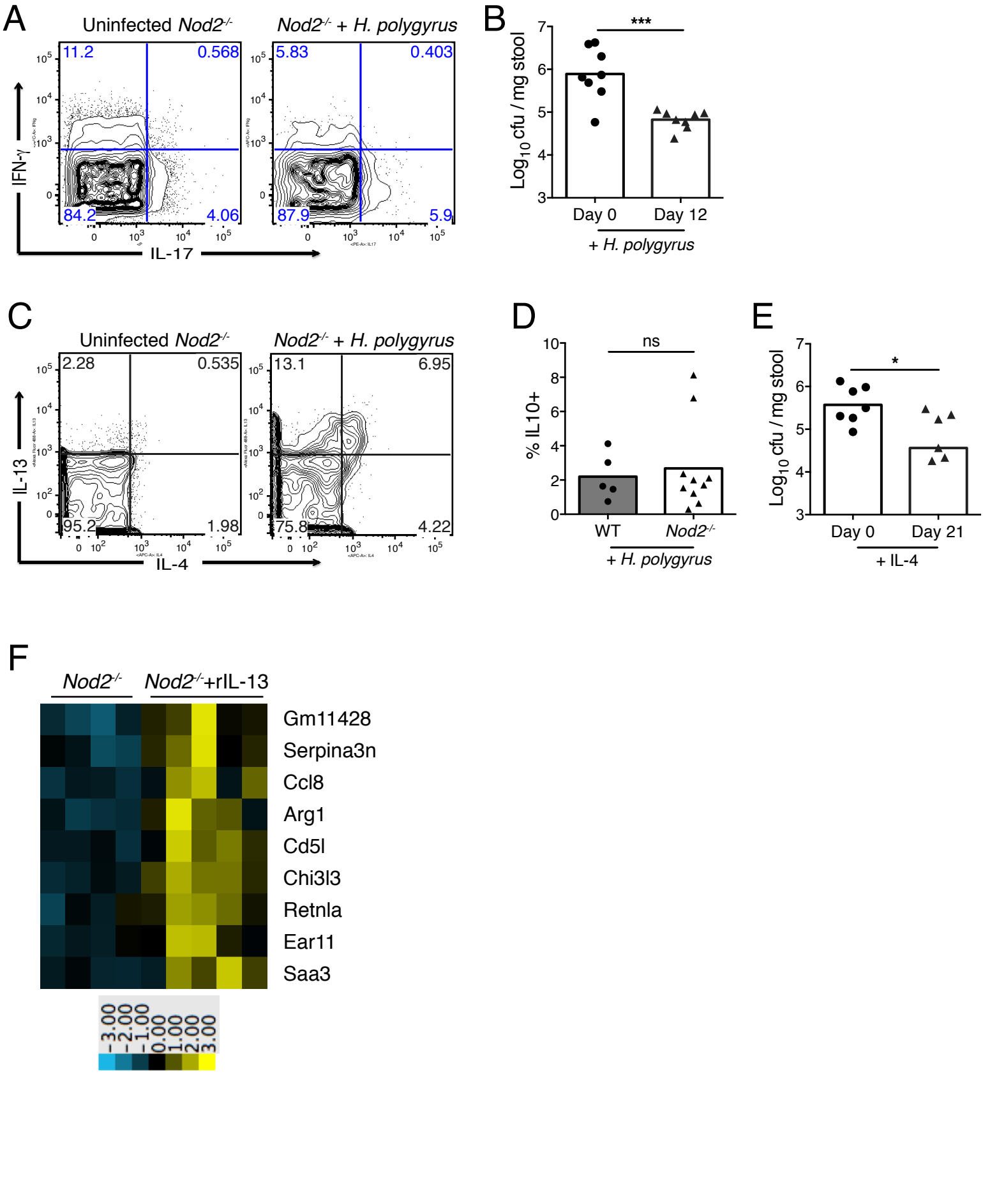


Figure S6: Helminth mediated protection from intestinal abnormalities is associated with the type-2 response.

(A) Representative flow cytometry plot of $CD8^+$ IELs in uninfected and *H. polygyrus* infected WT and *Nod2*$^{-/-}$ mice stained for intracellular IFN-γ and IL-17 expression after stimulation with PMA and ionomycin gated on $CD3^+$ live cells (n=6 mice per genotype). **(B)** Quantification of *B. vulgatus* in stool of *H. polygyrus* infected *Nod2*$^{-/-}$ mice 12 days post infection (n=8 mice per genotype). The partial reduction of *B. vulgatus* in stool compared to *T. muris* is probably because *H. polygyrus* and *T. muris* occupy different niches in the intestine, small intestine and cecum/colon, respectively. **(C)** Representative flow cytometry plots of $CD4^+$ T cells in the small intestinal lamina propria of uninfected and *H. polygyrus* infected *Nod2*$^{-/-}$ mice stained for intracellular IL-13 and IL-4 expression after stimulation with PMA and ionomycin gated on $CD3^+$ live cells (n≥6 mice per genotype). *H. polygyrus* induced a greater increase in IL-13^+ CD4 T cells in the small intestinal LP of WT and *Nod2*$^{-/-}$ mice compared with *T. muris* (Figure 2D), which likely explains the large reduction of tissue-associated *B. vulgatus* burden following infection by this parasite (Figure 2F). **(D)** Proportion of $CD4^+$ T cells expressing IL-10 in the small intestinal lamina propria of *H. polygyrus* infected WT and *Nod2*$^{-/-}$ mice after stimulation with PMA and ionomycin 12 days post infection (n≥5 mice per genotype). In contrast to *T. muris* infection, there were no differences in IL-10 producing $CD4^+$ T cells between *H. polygyrus* infected WT and *Nod2*$^{-/-}$ mice. **(E)** Quantification of *B. vulgatus* in stool of *Nod2*$^{-/-}$ mice treated with recombinant IL-4 complex (n≥6 per genotype). **(F)** Heat map of the 9 genes that were upregulated in *Nod2*$^{-/-}$ mice treated with recombinant IL-13 compared to PBS controls. All 9 are associated with M2 macrophages. *$p<0.05$, and ***$p<0.001$ by paired t-test for (B) and (E), and unpaired t-test for (D). Data are representative of at least two independent experiments. Each data point represents an individual mouse, bar denotes mean.

Supplementary figure 7

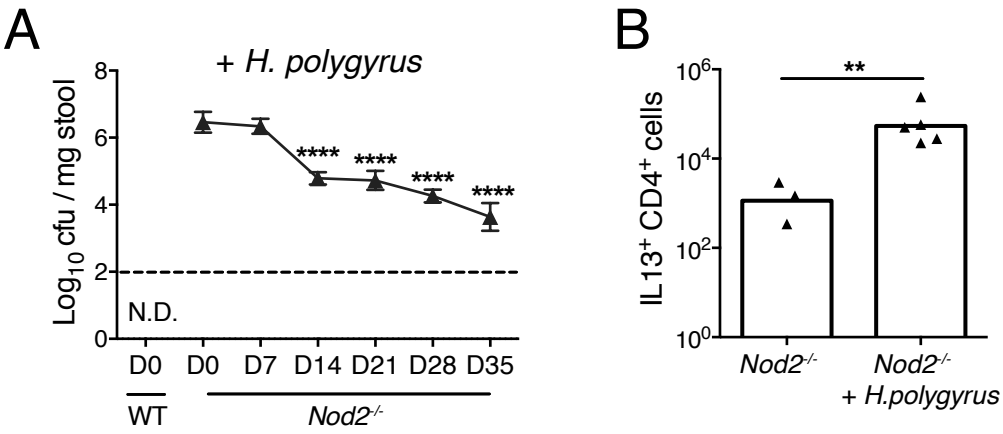


Figure S7: Chronic *H. polygyrus* infection leads to type-2 immunity mediated clearance of *B. vulgatus*.

(A) Longitudinal quantification of *B. vulgatus* in stool of *H. polygyrus* infected *Nod2*$^{-/-}$ mice (n=4 mice per genotype). **(B)** Quantification of CD4^{+} T cells expressing IL-13 in the small intestinal lamina propria of uninfected and *H. polygyrus* infected *Nod2*$^{-/-}$ mice after stimulation with PMA and ionomycin 35 days post infection (n≥3 mice per genotype). **p<0.01, and ****p<0.0001 by ANOVA with Holm-Sidak multiple comparisons test for (A), and unpaired t-test for (B). Data are representative of at least two independent experiments. Data are represented as mean ± SEM in (A). Each data point represents an individual mouse, bar denotes mean in (B).

Supplementary figure 8

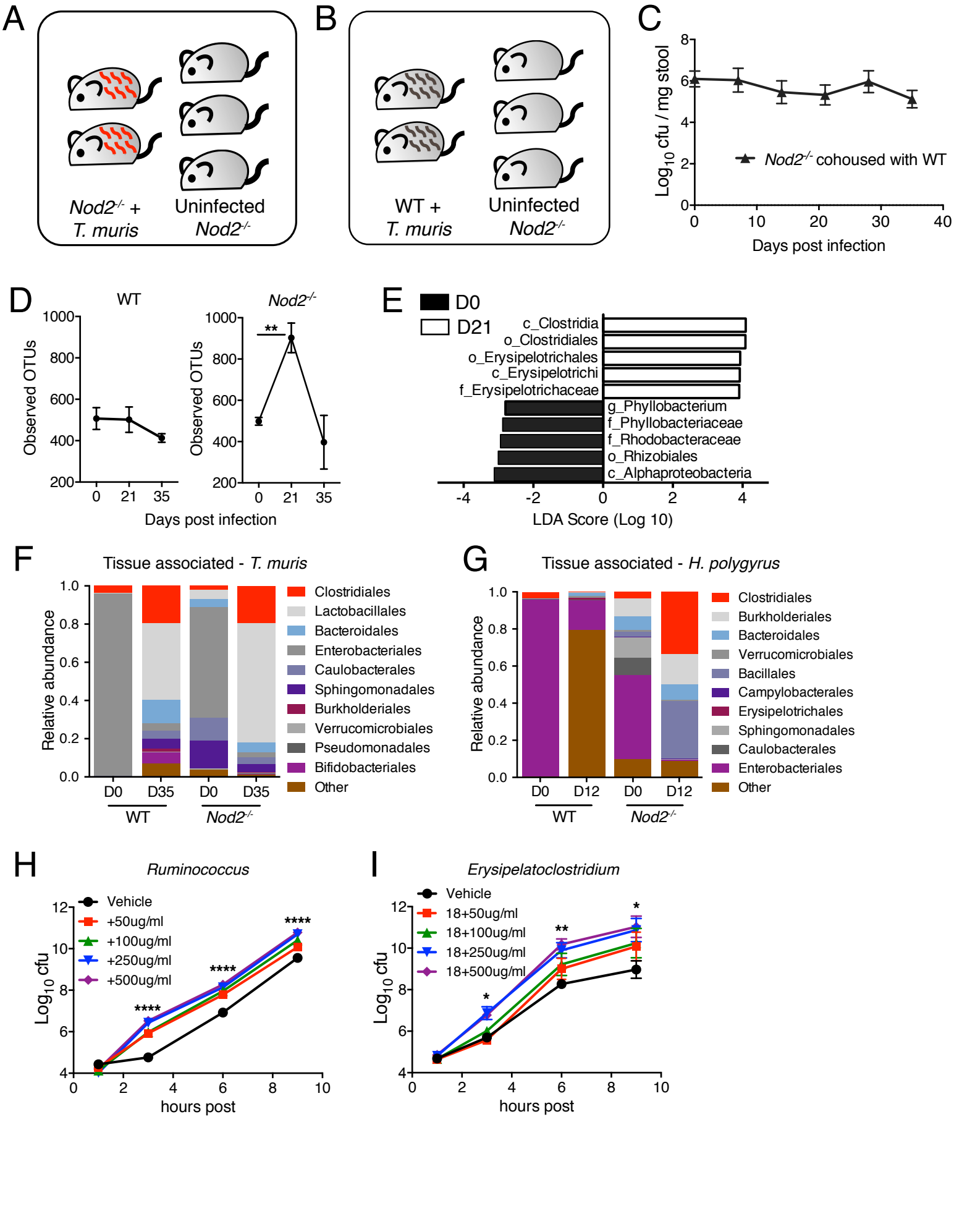
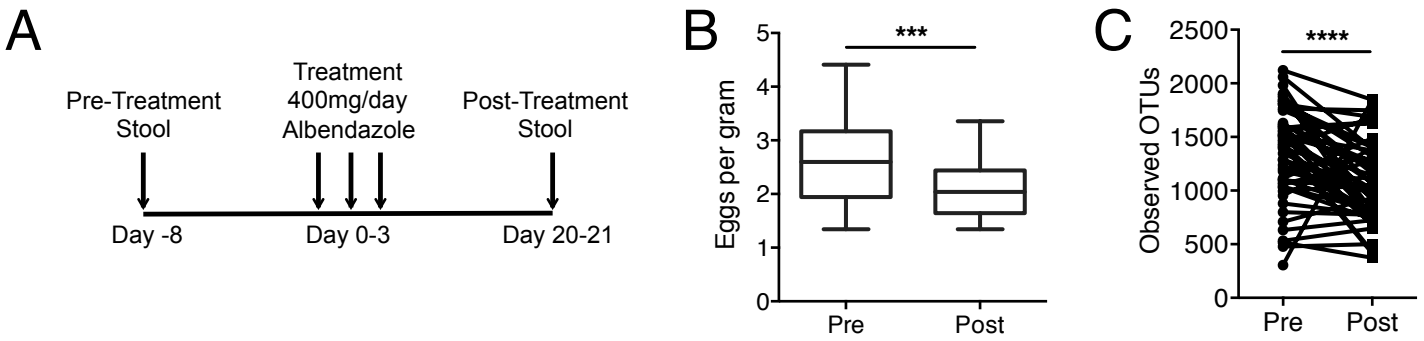
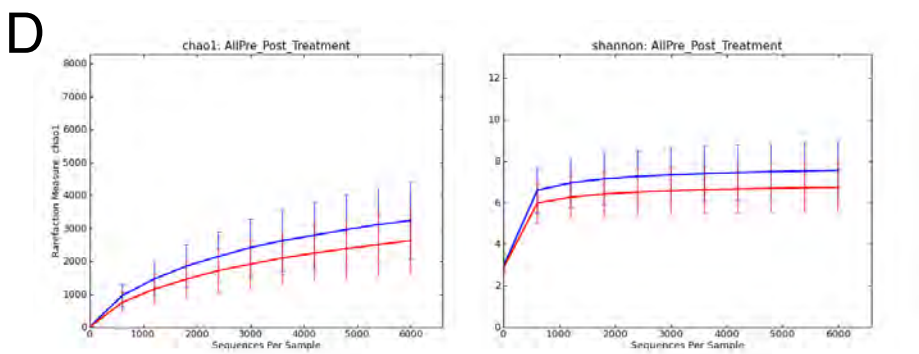


Figure S8: Alterations to the microbiota after helminth infection are distinct for WT and *Nod2*$^{-/-}$ mice. (A-B) Schematic for co-housing uninfected *Nod2*$^{-/-}$ and *T. muris* infected *Nod2*$^{-/-}$ mice (A), and uninfected *Nod2*$^{-/-}$ and *T. muris* infected WT mice (B). **(C)** Quantification of *B. vulgatus* in stool harvested from uninfected *Nod2*$^{-/-}$ mice co-housed with *T. muris* infected WT mice as in (A) (n=3). Unlike co-housing with *T. muris*-infected *Nod2*$^{-/-}$ mice (Fig. 3A), co-housing with *T. muris*-infected WT mice does not lead to a reduction in *B. vulgatus* burden in naïve *Nod2*$^{-/-}$ mice. **(D)** Alpha-diversity in *T. muris* infected WT and *Nod2*$^{-/-}$ stool at day 0, 21, and 35 post *T. muris* infection. (n≥5 per genotype). **(E)** LEfSE analysis to determine alterations to the stool microbiota of *Nod2*$^{-/-}$ mice after recombinant IL-4 treatment using an LDA threshold score of 3. (n≥6 per genotype). **(F)** Relative abundance of taxonomic groups in response to *T. muris* infection in the small intestinal tissue of WT and *Nod2*$^{-/-}$ mice as determined by 16S sequencing (n≥5 per genotype). **(G)** Relative abundance of taxonomic groups in response to *H. polygyrus* infection in the small intestinal tissue of WT and *Nod2*$^{-/-}$ mice as determined by 16S sequencing (n≥5 per genotype). A robust expansion of Clostridia is observed with both helminth infections, consistent with 16S analysis of stool. **(H-I)** Quantification of *Ruminococcus* species (Clostridia mixture species #13) (H) or *Erysipelatoclostridium species* (Clostridia mixture species #18) (I) in the presence of varying concentrations of pig intestinal mucin or vehicle in the culture media. **p<0.01, ****p<0.0001 by ANOVA with Holm-Sidak multiple comparisons test for (D), (H), and (I). Asterisks represent p-value between vehicle control and all concentrations of mucin in (H) and (I). Data are represented as mean ± SEM from at least two independent experiments.

Supplementary figure 9





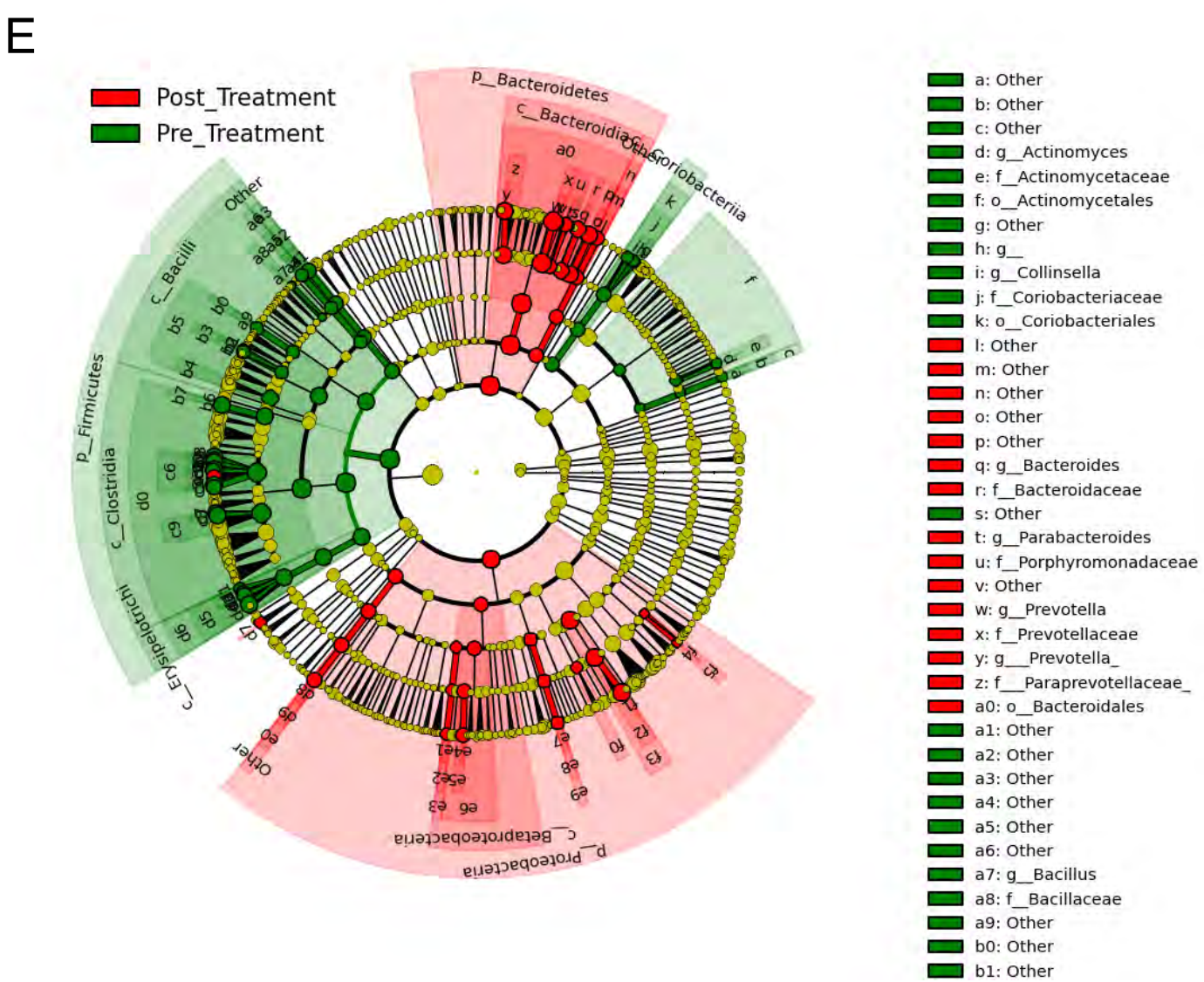


Figure S9: Albendazole treatment of the Orang Asli reduces microbial diversity of the gut microbiota and alters the composition of bacterial communities. (A) Schematic of the sample collection and deworming regimen for the longitudinal study of the microbiota of Orang Asli. **(B)** *Trichuris trichiura* egg burden is reduced post treatment with Albendazole (3 doses, 400mg/day) despite low cure rates (32.9%). Box and whisker plot shows alteration in eggs per gram of feces as determined by traditional Kato-Katz method. **(C)** Reduced alpha-diversity (as Observed OTUs) post treatment with Albendazole in paired analyses. **(D)** Rarefaction curves calculated for the Chao 1 index (Left) ($p=0.009$) and Shannon index (Right) ($p=0.003$) estimators of alpha diversity show reduced microbial diversity post treatment (red lines) compared to pre treatment (blue lines). **(E)** Visualization using a cladogram of bacterial taxa identified to be significantly different between pre-treatment and post-treatment stool samples of the Orang Asli utilizing LEfSe with a log LDA score above 3.00. ***$p<0.001$, ****$p<0.0001$ by paired t-test for (B) and (C).

Supplementary figure 10

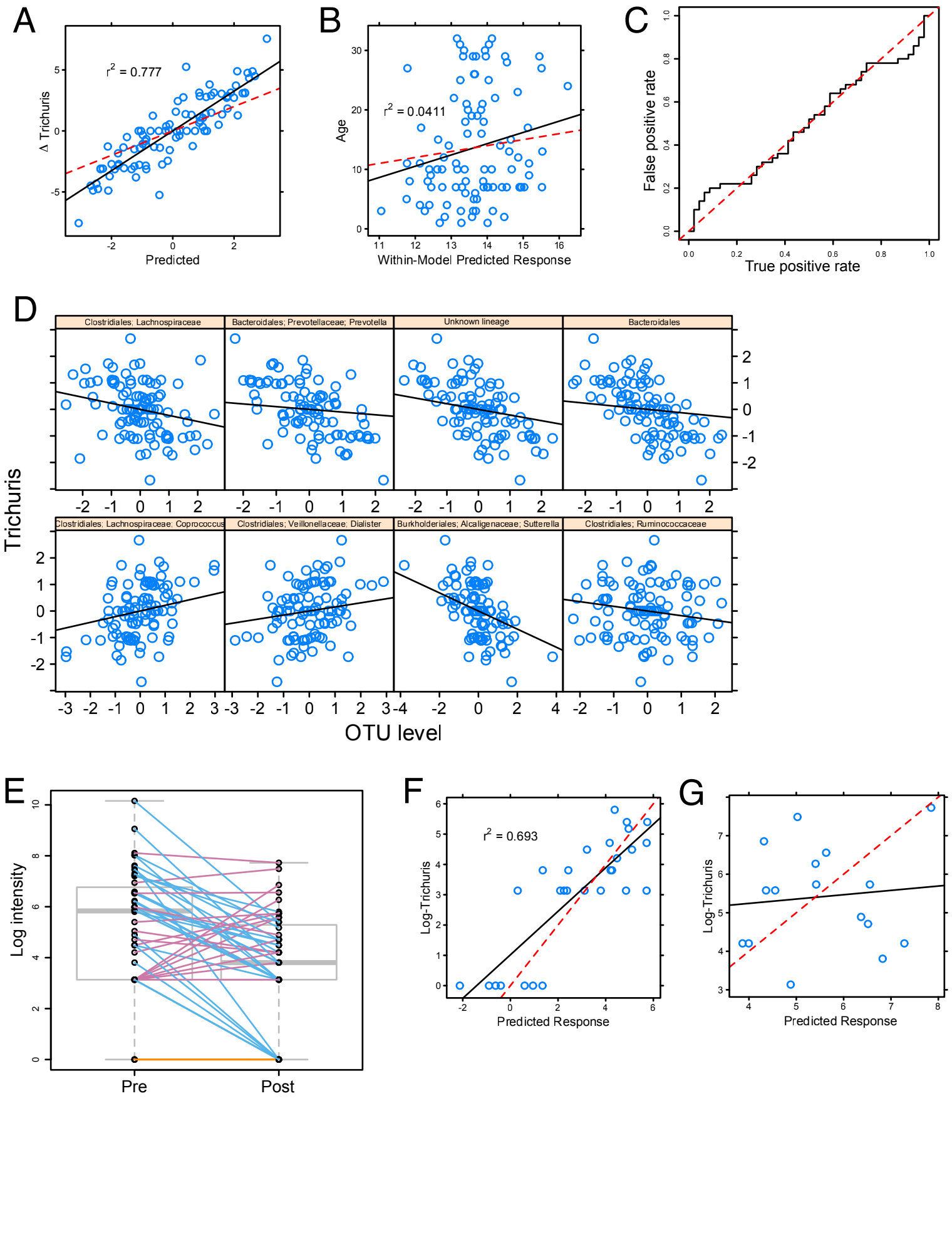


Figure S10: The effects of Albendazole on the gut microbiota of the Orang Asli is associated with *Trichuris trichiura* egg burden. (A-C) Centered log-ratio (clr) transformation of microbial compositions (from 16S sequencing data) followed by spares Partial Least Squares regression (sPLS) analyses was used to compute a model on within-subject changes for clr-transformed microbial community compositions and (log-transformed) *Trichuris trichiura* egg burdens (A). Change in *Trichuris trichiura* egg burden is strongly associated with a minimal set of taxa and the linear combinations of selected microbial taxa showed accurate prediction (r2=0.777) of change *Trichuris trichiura* egg burdens (ΔTrichuris) and no predictive value for Age (B). Scatter plots A and B illustrate the relationship between change in egg burdens and Age with the within model predicted responses from the sPLS regression analyses. There is also predicted relationship with Gender (C), where a receiver operator characteristic (ROC) curve showing the true positive vs. false positive rates over predicting various characteristics points to essentially random predictions at AUC = 0.5 (and follows closely the x=y line), whereas perfect prediction would have an AUC=1. Hence, selected taxa were found to accurately predict change in *Trichuris trichiura* egg burdens (ΔTrichuris), but not Age or Sex despite being included in the model. **(D)** Scatterplots illustrating the relationship between specific OTUs identified as the minimal set of taxa predictive of changes in *Trichuris trichiura* egg burdens. **(E)** To separate *T. trichiura* into "responders" (Blue lines) and "non-responders" (Pink lines) to Albendazole treatment, a cutoff of - delta 0.03 log egg burden was used (requiring the presence of *T. trichiura* before treatment). **(F-G)** As expected, sPLS-regression accurately associated specific bacterial taxa abundance with changes in *T. trichiura* burden in the "responders" (R2=0.693) (F) but not in "non-responders" (G). These results show that the effects of *T. trichiura* infection on the microbial communities described are directly driven by changes in worm burden, independently of Age, Gender, and Albendazole treatment.

Supplementary figure 11

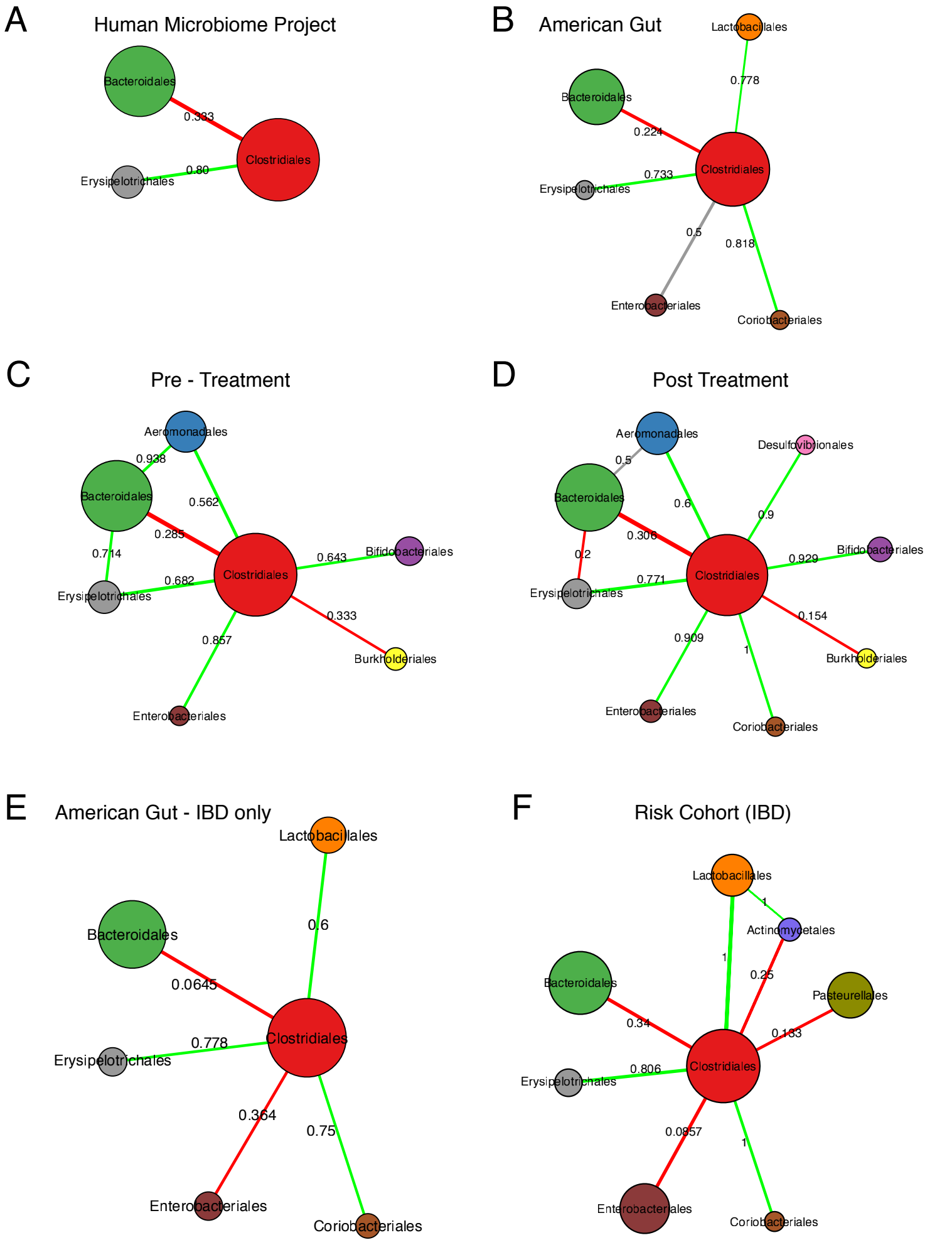


Figure S11: Inferring microbial networks by SPIEC-EASI reveals an antagonistic relationship between Clostridiales and Bacteroidales. (A-F) SPIEC-EASI (SParse InversE Covariance Estimation for Ecological Association Inference) is a statistical method for the inference of microbial ecological networks. This approach addresses the problems of compositional and sparse datasets inherent in 16S based investigations of microbial communities (see Supplemental Methods). Application of this method in an unbiased fashion to the gut microbiome data from the Human Microbiome Project (A) and the American Gut Project (B) identified a consistent negative association (red line) between Bacteroidales and Clostridiales, as well as a positive association (green line) between Erysipelotrichales and Clostridiales. Networks are visualized with OTU nodes colored by Order, with the node diameter proportional to the geometric mean of the OTU's relative abundance. Numerical values on the edges are the fraction of edges that are either majority positive (Green) or majority negative (Red). In one case there are exactly equal numbers of positive and negative examples (Grey). These positive and negative relationships were also observed in the gut microbiota data from the Orang Asli before deworming treatment (C) and after deworming treatment (D). Among IBD patients in the American Gut Project (E) and the Pediatric RISK Cohort (F), there is also a strong positive association between Clostridiales and Lactobacillales, which is not observed among the Orang Asli. Whereas there is a strong positive association between Clostridiales and Enterobacteriales among the Orang Asli (A & B), this becomes a negative association among IBD patients in the USA that are part of the American Gut Project (E) or the RISK Cohort (F). These results indicate that the antagonistic relationship between Clostridiales and Bacteroidales that we identified with helminth infection mouse models and a small dataset of helminth infected individuals may reflect a general phenomenon reflected in much larger datasets of USA residents. This negative relationship is the most consistently observed negative relationship in all available datasets.

Supplementary table 1

RefSeq	Log2 Fold Change	p-value	Gene Symbol	Description
NM_001081957	1.530004201	3.48E-05	Gm11428	activated macrophage/microglia WAP domain
NM_007482	1.46752967	1.92E-05	Arg1	arginase-1
NM_009252	1.580151604	9.40E-06	Serpina3n	serine protease inhibitor A3N precursor
NM_009690	1.504428477	8.47E-05	Cd5l	CD5 antigen-like precursor
NM_009892	2.680582506	3.05E-12	Chi3l3	chitinase-3-like protein 3 precursor
NM_011315	1.846496689	6.94E-07	Saa3	serum amyloid A-3 protein precursor
NM_020509	1.625305541	2.67E-05	Retnla	resistin-like alpha
NM_021443	1.499722438	8.21E-05	Ccl8	C-C motif chemokine 8 precursor
NM_053113	1.697677309	5.51E-06	Ear11	non-secretory ribonuclease

Table S1: *Nod2*$^{-/-}$ mice treated with rIL-13 display an increase in M2 macrophage activation genes.

List of genes upregulated in *Nod2*$^{-/-}$ mice treated with recombinant IL-13 compared to PBS controls.

Supplementary table 2

	Pre-Treatment (N=75)			Urban individuals (N=19)		
	Positive samples	**Prevalence (%)**	**95% CI**	**Positive samples**	**Prevalence (%)**	**95% CI**
Intestinal helminths	72	96.0	91.57-100.43	1	5.3	-4.80-15.40
T. trichiura	70	93.3	87.68-98.98	-	-	-
Hookworm	27	36.0	16.95-37.05	1	5.3	-4.8-15.40
Ascaris spp.	10	13.3	5.64-21.02	-	-	-

Table S2: Prevalence of intestinal helminth infections among Orang Asli and urban individuals.

Table representing the incidence of different intestinal helminths among the rural Orang Asli individuals before deworming treatment, and urban individuals residing in Kuala Lumpur.

Supplementary table 3

	T. trichiura		*Ascaris* spp.		Hookworm	
Infection intensity (%)						
Before Treatment (range of epg)	22-25,718 epg[b]		22-21,824 epg		22-1,999 epg	
Light-intensity infections	50	66.7	7	9.3	27	36.0
Moderate-intensity infections	19	25.3	3	4.0	-	-
Heavy-intensity infections	1	1.3	-	-	-	-
Total	**70**	**93.3**	**10**	**13.3**	**27**	**36.0**
After Treatment	22-2,266 epg		22-4999 epg		22	
Light-intensity infections	45	70.3	2	3.1	1	1.6
Moderate-intensity infections	2	3.1	-	-	-	-
Total	**47**	**73.4**	**2**	**3.1**	**1**	**1.6**
Cure rate (%)[a]						
20 days	23	32.9	8	80.0	26	96.3

a: Number of infected person cured after deworming/number of infected person before deworming

b: Eggs per gram of faecal sample

Table S3: Efficacy of albendazole in treating intestinal helminth infections according to type of parasites among Orang Asli in Kuala Pangsun.

Table representing de-worming efficacy in the Orang Asli population pre-treatment (n=75) and post-treatment (n=64).

Supplementary table 4

Tax ID	Phylum	Taxa	Trichuris (p-value, sign)	Age	Sex
OTU684		Bacteria	1.333e-04 (-)	-	-
OTU927	Bacteroidetes	Prevotella	7.467e-03 (-)	-	-
OTU1185	Bacteroidetes	Bacteroidales	8.400e-03 (-)	-	-
111135	Proteobacteria	Sutterella	<3.333e-05 (-)	-	-
317814	Firmicutes	Ruminococcaceae	8.000e-04 (-)	-	-
4414476	Firmicutes	Lachnospira	1.067e-03 (-)	-	-
182289	Firmicutes	Coprococcus	4.000e-04 (+)	-	-
174516	Firmicutes	Dialister	6.734e-03 (+)	-	-

Table S4: Taxa selected by multi-level sPLS in response to deworming.

Taking advantage of the repeated measures design, we used a sPLS model on within-subject variances of clr- transformed compositions and (log-transformed) *Trichuris* burden we found that the change in *Trichuris* burden is strongly associated with a minimal set of taxa. The resulting models show a clear positive association for intensity of *T. trichiura* infection with a negative association with Bacteroidia populations, *Prevotella* and Bacteroidales. Two Clostridia OTUs show a clear positive association, *Coprococcus* and *Dialister*, with Ruminococcaceae and Lachnospira showing the opposite association that is nevertheless orthogonal to Bacteroidia. While linear combinations of these selected taxa show robust within-model error (∆*Trichuris*), no selected taxa was found to predict Age or Sex despite being included in the model. We classified patients that that had under deworming as either responding or not responding to *Trichuris* deworming (using a cut off of -∆0.03 log-*Trichuris*) and requiring the presence of *Trichuris* before deworming. We found that sPLS, regressing on log *Trichuris* burden was accurate at explaining worm burder (r^2 = 0.693) but was not predictive of *Trichuris* levels in non-responders.